UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
09/607,710	06/30/2000	Marc Alexander Najork	18449772-0288-999	8784

7590 09/09/2004

GARY S. WILLIAMS
PENNIE & EDMONDS LLP
3300 HILLVIEW AVENUE
PALO ALTO, CA 94304

EXAMINER	
OSMAN, RAMY M	

ART UNIT	PAPER NUMBER
2157	

DATE MAILED: 09/09/2004

Please find below and/or attached an Office communication concerning this application or proceeding.

Office Action Summary	Application No.	Applicant(s)	
	09/607,710	NAJORK ET AL.	
	Examiner	Art Unit	
	Ramy M Osman	2157	

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTH(S) FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If the period for reply specified above is less than thirty (30) days, a reply within the statutory minimum of thirty (30) days will be considered timely.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

1)☒ Responsive to communication(s) filed on *24 May 2004*.

2a)☒ This action is **FINAL**. 2b)☐ This action is non-final.

3)☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

4)☒ Claim(s) *1-52* is/are pending in the application.

4a) Of the above claim(s) _____ is/are withdrawn from consideration.

5)☐ Claim(s) _____ is/are allowed.

6)☒ Claim(s) *1-52* is/are rejected.

7)☐ Claim(s) _____ is/are objected to.

8)☐ Claim(s) _____ are subject to restriction and/or election requirement.

Application Papers

9)☐ The specification is objected to by the Examiner.

10)☐ The drawing(s) filed on _____ is/are: a)☐ accepted or b)☐ objected to by the Examiner.

Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).

Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).

11)☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

Priority under 35 U.S.C. § 119

12)☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).

a)☐ All b)☐ Some * c)☐ None of:

1.☐ Certified copies of the priority documents have been received.

2.☐ Certified copies of the priority documents have been received in Application No. _____.

3.☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

* See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

1) ☒ Notice of References Cited (PTO-892)
2) ☐ Notice of Draftsperson's Patent Drawing Review (PTO-948)
3) ☐ Information Disclosure Statement(s) (PTO-1449 or PTO/SB/08) Paper No(s)/Mail Date _____.
4) ☐ Interview Summary (PTO-413) Paper No(s)/Mail Date. _____ .
5) ☐ Notice of Informal Patent Application (PTO-152)
6) ☐ Other: _____.

DETAILED ACTION

Status of Claims

1. This communication is responsive to the amendment filed on May 24, 2004. Claims 1,13,22,23,31,38 and 46 were amended. Claims 1-52 are pending. The rejection are as stated below.

Claim Rejections - 35 USC § 103

2. The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

> (a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negatived by the manner in which the invention was made.

3. Claims 1-4,7-10,13-17,20-25,28-33,36-40,43-48,51 and 52 rejected under 35 U.S.C. 103(a) as being obvious over Monier (U.S. Patent No. 5,974,455) in view of Najork (U.S. Patent No. 6,301,614).

4. In reference to claims 1,13,23,31,38 and 46, Monier teaches downloading data sets from among a plurality of host computers comprising the following steps:

Storing representations of data set addresses in a set of data structures, including a first buffer, a second buffer and a first disk file, wherein representations of data set addresses stored

in the first disk file are ordered (column 3, lines 1-35, Monier discloses storing URL representations in a set of data structures, including a hash table (stored in random access memory (RAM)), an append buffer (stored in RAM) and a sequential disk file, wherein the representations are stored sequentially in the disk file).

Selecting as a current buffer one of the first and second buffers (column 6, lines 35-45, Monier discloses selecting and managing a current buffer among the hash table and append buffer).

Downloading at least one data set that includes addresses of one or more referred data sets (column 5, lines 20-30, Monier discloses fetching web pages that include URL's of one or more referred web pages).

Identifying the addresses of the one or more referred data sets (column 5, lines 20-30, Monier discloses analyzing and identifying the addresses of the one or more referred web pages).

For each identified address:

Generating a representation of the identified address (column 5 line 55 – column 6 line 22, Monier discloses generating a fingerprint representation of the specified URL), and

Determining whether the representation is stored in the buffer without determining whether the representation is stored in the first disk file, and when this determination is negative, storing the representation in the buffer (column 5 line 43 – column 6 line 22, Monier discloses determining whether the representation is stored in the hash table, and when this determination is negative, storing the representation in the hash table). Monier inherently teaches without determining whether the representation is stored in the first disk file.

When the buffer reaches a predefined full condition:

Ordering the contents of the buffer according to the representations (column 6, lines 1-33, Monier discloses ordering the contents of the hash table according to the fingerprint representations), and

Monier discloses appending contents of the hash table into the contents of the disk file (column 6 line 22 – column 7 line 12). Monier fails to teach performing an ordered merge of the contents of the buffer into the contents of the first disk file wherein the ordered merge comprises preventing duplication of any of the representations of data set addresses stored in the first disk file. However, Najork teaches sorting an index of representations and performing a sorted merge of the index with a disk file (column 3 line 30 – column 4 line 45 and column 6 line 1 – column 7 line 25).

It would have been obvious for one having ordinary skill in the art to perform a sorted merge of the hash table with the disk file as per the teachings of Najork for facilitating look-up operations on the disk file.

Selecting the other buffer as the current buffer, wherein the previously current buffer is identified as a non-current buffer (column 6, lines 22-67, Monier discloses selecting the append buffer as the current buffer, wherein the hash table is identified as a non-current buffer).

5. In reference to claims 2,14,24,32,39 and 47, Monier in view of Najork teach the method, the computer program and the web crawler system of claims 1,13,23,31,38 and 46 above, wherein after determining that the representation is not stored in the buffer, the identified address is stored in the buffer (column 5 line 43 – column 6 line 22, Monier discloses that after

determining that the representation is not stored in the hash table, the identified address is stored in the hash table).

6. In reference to claims 3,15,25,33,40,48, Monier in view of Najork teach the method, the computer program and the web crawler system of claims 1,13,23,31,38 and 46 above, wherein:

After determining that the representation is not stored in the buffer, the identified address is stored in a second disk file (column 9, lines 25-40, Monier discloses after determining that the representation is not stored in the hash table, the identified address is stored in a second disk file), and

Additionally storing with each representation in the buffer a pointer to the corresponding address stored in the second disk file (column 3, lines 1-20, column 5 & column 6, lines 20-53, Monier discloses additionally storing with each representation in the RAM a pointer to the corresponding address stored in the second disk file), and

While ordering the contents of the buffer, keeping with each representation in the buffer its pointer to the corresponding address in the second disk file (column 5 & column 6, lines 20-53, Monier discloses while ordering the contents of the hash table (in RAM), keeping with each representation in the hash table its pointer to the corresponding address in the disk file).

7. In reference to claims 4 and 16, Monier in view of Najork teach the method of claims 3 and 15 above, wherein when the buffer reaches a predefined full condition:

Each representation in the buffer stores an associated flag, setting the flag to a first value when the representation is equal to a representation previously stored in the first disk file, and setting the flag to a second value, when the representation is not equal to any representation previously stored in the first disk file (column 5 lines 25-35, & column 8 lines 45-65, Monier

discloses each representation in the hash table stores an associated " fetched flag", setting the flag to a first value when the representation is equal to a representation previously stored in the disk file, and setting the flag to a second value, when the representation is not equal to any representation previously stored in the disk file), and

Each representation whose flag is set to the second value, scheduling the corresponding data set for downloading (column 9, lines 25-50, Monier discloses each representation whose flag is set to the second value and marked as "not fetched", scheduling the corresponding data set for fetching).

8. In reference to claims 7,20,28,36,43 and 51, Monier in view of Najork teach the method, the computer program and the web crawler system of claims 1,13,23,31,38 and 46 above, wherein the representation of the identified address comprises a checksum of at least a portion of the identified address (column 5 line 55 – column 6 line 22, Monier discloses the representation of the identified URL comprising a fingerprint of at least a portion of the identified URL).

9. In reference to claims 8,21,29 and 44, Monier in view of Najork teach the method, the computer program and the web crawler system of claims 1,13,23 and 38 above, wherein: Determining whether the representation is stored in a cache before determining whether the representation is stored in the buffer (columns 6&7, Monier discloses determining whether the representation is stored in append buffer before determining whether the representation is stored in an input buffer (in RAM)), and

Determining whether the representation is stored in a cache, and if positive, skipping the determination of whether the representation is stored in the buffer (columns 6&7, Monier

discloses determining whether the representation is stored in an append buffer, and if positive, skipping the determination of whether the representation is stored in the input buffer), and

When the representation is not stored in the cache, the cache has not reached a predefined full condition, and other predefined criteria are met, adding the representation to the cache (columns 6&7, Monier discloses when the representation is not stored in the append buffer, the host name table has not reached a predefined full condition, and other predefined criteria are met, adding the representation to the input buffer), and

When the representation is not stored in the cache, the cache has reached said predefined full condition, and said other predefined criteria are met, evicting a stored representation from the cache in accordance with an eviction policy and adding the representation to the cache (columns 6&7, Monier discloses when the representation is not stored in the append buffer, the append buffer has reached said predefined full condition, and said other predefined criteria are met, evicting a stored representation from the append buffer in accordance with an eviction policy and adding the representation to the append buffer).

10. In reference to claims 9,10,17,30,37,45 and 52, Monier in view of Najork teach the method, the computer program and the web crawler system of claims 1,23,31,38 and 46 above, wherein when a representation in the first buffer is not found in the first disk file during merging, scheduling the corresponding data set for downloading (columns 6-8, Monier discloses that when a representation in the hash table is not found in the disk file during merging, scheduling the corresponding web page for fetching).

11. Claim 22 rejected under 35 U.S.C. 103(a) as being obvious over Monier (U.S. Patent No. 5,974,455) in view of Najork (U.S. Patent No. 6,321,265) in further view of Najork (U.S. Patent No. 6,301,614).

12. Monier teaches downloading data sets from among a plurality of host computers comprising the following steps:

Storing representations of data set addresses in a set of data structures, including a first buffer, a second buffer and a first disk file, wherein representations of data set addresses stored in the first disk file are ordered (column 3, lines 1-35, Monier discloses storing URL representations in a set of data structures, including a hash table (stored in random access memory (RAM)), an append buffer (stored in RAM) and a sequential disk file, wherein the representations are stored sequentially in the disk file).

Selecting as a current buffer one of the first and second buffers (column 6, lines 35-45, Monier discloses selecting and managing a current buffer among the hash table and append buffer).

Downloading at least one data set that includes addresses of one or more referred data sets (column 5, lines 20-30, Monier discloses fetching web pages that include URL's of one or more referred web pages).

Identifying the addresses of the one or more referred data sets (column 5, lines 20-30, Monier discloses analyzing and identifying the addresses of the one or more referred web pages).

Generating a representation of the identified address (column 5 line 55 – column 6 line 22, Monier discloses generating a fingerprint representation of the specified URL), and

Monier discloses determining whether the representation is stored in the hash table (column 5 line 43 – column 6 line 22). Monier fails to teach whether the disk file is empty, and when the representation is not stored in the buffer and the disk file is empty, scheduling the corresponding data set for downloading. However, Najork '265 teaches determining if a queue is empty and if it is empty then downloading data set addresses to the queue (column 3 line 1 – column 4 line 5).

It would have been obvious for one of ordinary skill in the art to download the data set corresponding to the representations in the hash table to the disk file if the disk file is empty as per the teachings of Najork '265 so that new URLs can be stored as they are processed.

Monier discloses determining if the representations have been previously stored in the hash table/disk file (columns 8&9). Monier fails to teach when the representation is not stored in the buffer and the disk file is not empty, storing the representation in the buffer and delaying scheduling of the corresponding data set for downloading until it is determined that the representation has not been previously stored in the disk file. However, Najork '265 teaches determining if the queue is not empty then delaying and assigning a download time for the data set addresses (column 3 line 1 – column 4 line 5).

It would have been obvious for one ordinarily skilled in the art to assign a download time for the data set addresses as per the teachings of Najork '265 that would allow sufficient time to determine if the representations are stored in the disk file.

Monier fails to teach performing an ordered merge of the contents of the buffer into the contents of the first disk file wherein the ordered merge comprises preventing duplication of any of the representations of data set addresses stored in the first disk file. However, Najork '614

teaches sorting an index of representations and performing a sorted merge of the index with a disk file, preventing duplication (column 3 line 30 – column 4 line 45 and column 6 line 1 – column 7 line 25).

It would have been obvious for one having ordinary skill in the art to perform a sorted merge of the hash table with the disk file as per the teachings of Najork '614 for preventing duplication and facilitating look-up operations on the disk file.

13. Claims 5,6,11,12,18,19,26,27,34,35,31,32,49 and 50 are rejected under 35 U.S.C. 103(a) as being unpatentable over Monier (U.S. Patent No. 5,974,455) in view of Najork (U.S. Patent No. 6,301,614) in further view of Cabrera et al. (U.S. Patent No. 5,953,729).

14. In reference to claims 5,11,18,26,34,41 and 49, Monier teaches the method, the computer program and the web crawler system of claims 1,13,23,31,38 and 46 above.

Monier does not teach storing representations of data set addresses in a sparse disk file which is divided into portions (or sub-files), each portion having a starting address and contents comprising an ordered list of representations of data addresses. However, Cabrera teaches sparse file technology divided into clusters each having a cluster number (column 9, lines 40-66).

It would have been obvious to one having ordinary skill in the art to modify Monier by storing URL representations in a sparse file as per the teachings of Cabrera so as to minimize the overhead in managing and ordering the contents on the disk file.

15. Monier does not teach merging the contents of the buffer with the ordered contents of the sparse disk file to include determining a starting address for a corresponding portion of the

sparse disk file. However, Cabrera teaches sparse file technology which can indicate starting cluster numbers for portions of the sparse file (columns 9&10).

It would have been obvious to one having ordinary skill in the art to modify Monier by when merging the contents of the hash table with the ordered contents of the sparse file, to include determining a starting cluster number for a corresponding portion of the sparse disk file as per the teachings of Cabrera so as to minimize the overhead for merging and ordering of the contents on the disk file.

16. Monier does not teach merging the contents of the buffer with the ordered contents of the sparse disk file to include performing an ordered merge of a subset of the buffer, starting at the representation for which the starting address was obtained, into the contents of the corresponding portion. However, Cabrera teaches sparse file technology which can indicate starting cluster numbers for portions of the sparse file (columns 9&10).

It would have been obvious to one having ordinary skill in the art to modify Monier by when merging the contents of the hash table with the ordered contents of the sparse disk file to include performing an ordered merge of a subset of the hash table, starting at the representation for which the starting address was obtained, into the contents of the corresponding portion as per the teachings of Cabrera so as to minimize the overhead in merging and ordering the contents on the disk file.

17. In reference to claims 6,12,19,27,35,42 and 50, Monier teaches the method, the computer program and the web crawler system of claims 1,13,23,31,38 and 46 above.

18. Monier does not teach storing representations of data set addresses in a sparse disk file having empty entries interspersed among entries storing said representations. However, Cabrera

teaches sparse file technology which comprises a mixture of zero data and non-zero data (column 7, lines 20-50).

It would have been obvious to one having ordinary skill in the art to modify Monier by storing representations of data set addresses in a sparse disk file having zero data interspersed among data of said representations as per the teachings of Cabrera so as to minimize the overhead in sequentially ordering the data contents on the disk file.

19. Monier teaches sequentially scanning the disk file via an input buffer, starting at the representation for which a starting address was obtained, until a representation matching the respective representation is found (column 6 lines 35-67 & column 9 lines 25-50). Monier does not teach scanning the disk file until one of the empty entries is found, and when an empty entry is found storing the respective representation in the empty entry. However, Cabrera teaches sparse file technology which comprises a mixture of zero data and non-zero data (column 7, lines 20-50).

It would have been obvious to one having ordinary skill in the art to modify Monier by scanning the disk file until one of the zero data entries is found as per the teachings of Cabrera, and when zero data entry is found storing the respective representation in the zero data entry, so as to minimize the overhead of ordering the data contents on the disk file while merging the contents of the hash table with the contents of the disk file.

Response to Amendment

20. The examiner acknowledges amended claims 1,13,22,23,31,38 and 46 filed on 5/24/2004.

Response to Arguments

21. Applicant's arguments with respect to claims 1-52 have been fully considered but they are not persuasive.

22. Applicant argues that Najork '614 teaches preventing duplication before merging the cache, which is unlike embodiments of the present invention. However, Najork '614 teaches that the merging process involves the step of preventing duplication of any of the representations of data set addresses stored in the first disk file (column 3 line 30 – column 4 line 45 and column 6 line 1 – column 7 line 25). The claim language does not indicate that the step of preventing duplication occurs after merging, and therefore it is irrelevant as to when the prevention takes place.

23. In response to applicant's argument that the proposed combination is improper, it has been held that a prior art reference must either be in the field of applicant's endeavor or, if not, then be reasonably pertinent to the particular problem with which the applicant was concerned, in order to be relied upon as a basis for rejection of the claimed invention. See *In re Oetiker*, 977 F.2d 1443, 24 USPQ2d 1443 (Fed. Cir. 1992).

In this case, Monier teaches that the purpose of sequential ordering of disk entries is to eliminate random accesses to the disk, and thus minimize latency (column 3 lines 1-30 and column 6 line 20 – column 7 line 30). Therefore, the concept of sequentially appending the disk can be modified to incorporate the teachings of Najork '614. Najork '614, teaches that in order to

reduce random access to the disk, the entries are stored in sequential order. Najork '614 then teaches that after sorting in the mentioned order, a sorted merge is then performed between the index and disk. This is done for the purpose of facilitating subsequent look-up operations.

Therefore, it would have been obvious for one having ordinary skill in the art to perform a sorted merge of the hash table with the disk file as per the teachings of Najork for facilitating look-up operations on the disk file

Conclusion

24. Applicant's amendment necessitated the new ground(s) of rejection presented in this Office action. Accordingly, **THIS ACTION IS MADE FINAL**. See MPEP § 706.07(a). Applicant is reminded of the extension of time policy as set forth in 37 CFR 1.136(a).

A shortened statutory period for reply to this final action is set to expire THREE MONTHS from the mailing date of this action. In the event a first reply is filed within TWO MONTHS of the mailing date of this final action and the advisory action is not mailed until after the end of the THREE-MONTH shortened statutory period, then the shortened statutory period will expire on the date the advisory action is mailed, and any extension fee pursuant to 37 CFR 1.136(a) will be calculated from the mailing date of the advisory action. In no event, however, will the statutory period for reply expire later than SIX MONTHS from the date of this final action.

Any inquiry concerning this communication or earlier communications from the examiner should be directed to Ramy M Osman whose telephone number is (703) 305-8050. The examiner can normally be reached on M-F 9-5.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Ario Etienne can be reached on (703) 308-7562. The fax phone number for the organization where this application or proceeding is assigned is 703-872-9306.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see http://pair-direct.uspto.gov. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free).

RMO
September 3, 2004

ARIO ETIENNE
SUPERVISORY PATENT EXAMINER
TECHNOLOGY CENTER 2100